

EVALUATION MANUAL

EVALUATING OPERATIONS RESEARCH UTILIZATION:
GUIDELINES FOR ASSESSING PROCESS AND IMPACT

M. CELESTE MARIN
JANE T. BERTRAND

FRONTIERS IN REPRODUCTIVE HEALTH

POPULATION COUNCIL
FAMILY HEALTH INTERNATIONAL
TULANE UNIVERSITY

USAID COOPERATIVE AGREEMENT NO. HRN-A-00-98-00012-00

OVERVIEW

The purpose of this document is to describe the FRONTIERS evaluation methodology and give detailed instructions on its implementation. It is intended for use primarily by FRONTIERS project monitors, who will be conducting the evaluations. Below is a brief description of each section.

Part A: Concepts

Background

This section describes the need for a new approach to evaluation of FRONTIERS OR studies, how the methodology was developed and its main characteristics.

Implementation of the Evaluation Plan

This section describes the three main components of the evaluation methodology: 1) a process assessment by the prime monitor at the end of the subproject; 2) an impact assessment by the prime monitor two years later; and 3) verification visits by an external team to a subset of process and impact assessments, to confirm, supplement or refute findings. Limitations of the methodology are also included in this section.

Part B: Implementation

The Evaluation Procedure

This section describes the steps the project monitors should take to conduct the processor impact assessment. In addition to the project monitor's direct knowledge of the subproject, key informants and project documents are the two main sources of data to be used in the assessments.

The Assessment Form

The assessment form is a tool for both data collection and reporting. While the indicators differ between the process and impact assessments, both require the same combination of a numerical score and a narrative explanation for each indicator. This section describes how to complete the assessment forms; the forms themselves can be found in Attachments 1 and 2.

Indicators

This section lists the indicators and defines or explains each, including some examples and tips on scoring. Sample responses from previous OR subprojects accompany each indicator.

Evaluation Report

Every evaluation will be submitted to Celeste Marin in the FRONTIERS/DC office in the form of a summary report. Reports will consist primarily of the completed assessment form, but will also include some background details on the subproject, a description of how the evaluation was conducted and a section for open comments.

ATTACHMENTS

Attachment 1: Process Assessment Form

This is a blank Process Assessment Form that can be used as a template for both interviews and reporting.

Attachment 2: Impact Assessment Form

Similar to Attachment 1, this is a blank Impact Assessment Form.

PART A

CONCEPTS

I. BACKGROUND

Rationale for a New Evaluation System

The Population Council has been providing technical assistance and conducting Operations Research (OR) studies in family planning and reproductive health for over two decades. These studies are designed to identify problems and test strategies to resolve those problems, providing program managers, administrators and policymakers with the information they need to improve service delivery systems. The ultimate goal is that strategies tested through OR will be adopted, leading to changes in policy or operational procedures at institutional, national or international levels. These changes are what we refer to as impact. The purpose of the new approach to evaluation described in the document is to assess the magnitude of impact achieved by FRONTIERS, as well as the role of process and contextual factors.

Evaluation is nothing new to operations research. Under the USAID results framework, each study is designed to contribute to attainment of the strategic objective by achieving intermediate or lower level results. Most OR studies have an evaluation component to measure these results—in other words, whether a specific change in some aspect of service delivery was successful in accomplishing its objectives.

In addition, under the OR/TA regional agreements, external evaluations were carried out to assess the regional project on a broader scale, and case studies focusing on utilization were conducted in a number of countries or regions over the years. This combination of evaluation methods enabled the Population Council to assess the results of an intervention in the short term, or of groups of studies at a given point in time. However, the evaluation efforts required a large investment of time and money and ultimately they could not adequately answer the question “Did OR have a lasting impact on family planning and reproductive health service delivery?”

The Population Council and USAID recognized the need for more regular and systematic assessment to determine whether the FRONTIERS Strategic Objective (SO) of “Improved family planning and related reproductive health service delivery through OR” was accomplished. FRONTIERS was charged with developing an innovative approach to evaluating the process and impact for its portfolio of OR projects. Such an evaluation system would have numerous benefits.

- It would be possible to identify strengths and weaknesses of the project in an ongoing manner, rather than at a single point in time toward the end of the project.
- It could take advantage of the experience and expertise of FRONTIERS staff, resulting in more thorough findings than might be possible from a brief site visit by external evaluators.
- Lessons learned in one site could be applied in others, increasing the likelihood of successful interventions and utilization of findings.

- Data would be available to do other kinds of analyses, for example exploring the relationships between utilization and non-project variables such as family planning effort scores, or political or economic environment.

Principles Underlying the Evaluation Strategy

Over the course of the first two years, Tulane University developed an evaluation methodology for FRONTIERS conforming to the following principles:

- The information should be collected using a standard instrument across countries.
- The data should be formatted so they can be entered into the existing ACCESS database that tracks all FRONTIERS projects for the purposes of financial and technical monitoring, reporting to USAID, and other needs.
- The data collection instrument should provide both quantitative measurements of project impact and qualitative information that allows for a greater appreciation of the context in which the project was carried out.
- The instructions for completing these forms should be sufficiently straightforward for project monitors to be able to fill in the information on the forms on all OR projects upon completion and at a period of 24 months post-completion.

Two rounds of three case studies each were conducted to test and refine the methodology proposed by FRONTIERS/Tulane staff. The first round of case studies focused on identifying the types of changes to be defined as impact, while the second concentrated on refining the data collection instrument (referred to herein as the assessment form.) Each round of case studies included one country from each program region. The evaluation team assessed projects completed under the regional OR/TA agreements prior to 1998, collecting data on the indicators through key informant interviews and review of project documents. The methodology developed will be used for intervention and evaluative studies, but is not considered appropriate for diagnostic studies or technical assistance activities.

The assessment form consists of 14 process indicators, 11 impact indicators, and six contextual/other factors. Because the evaluation will be conducted in two distinct phases (unlike the case studies), the assessment form will be broken down into a Process Assessment Form and an Impact Assessment Form. The matrix format of the assessment forms has the following advantages:

- It provides a standardized format to the assessment of projects in each country.
- It combines quantitative and qualitative assessment by using a numerical score accompanied by justification for the score, for each indicator.
- It is appropriate both as a discussion guide for in-depth interviews with key informants and as a reporting format for subproject assessments.
- Data from multiple subprojects can be condensed into a single table or summary grid that provides an overview of the results for the entire set of studies over the full range of indicators.

II. IMPLEMENTATION OF THE EVALUATION PLAN

Data Collection

In general, FRONTIERS intervention and evaluation studies will be assessed by Population Council staff at two points in time:

1. Upon completion of the subproject, with the evaluation report submitted along with the subproject final report, and
2. 24 months after completion of the subproject, with a report submitted to the Washington, D.C. office two months later.

The first evaluation will concentrate primarily on the context and process of the study, as it will be too early to determine the full impact of the subproject on the policy and service delivery environment. After two years, sufficient time will have elapsed for an assessment of impact. All evaluations will involve: 1) completion of the relevant assessment form (either Process or Impact, described further in Section IV), and 2) submission of a summary report, described in Section VI, below.

The Associate Director for each region will review completed reports before they are sent to Washington, D.C. Both the Regional Associate Director and the Tulane Evaluation Associate will give the project monitor appropriate feedback on the evaluation, although there is no prescribed format for such feedback.

The purpose of the evaluation is to better understand the factors related to utilization of OR findings in order to increase the impact of FRONTIERS; thus project monitors should report on both positive and negative aspects of a study. Evaluations will not be used to assess the performance of the project monitor or any other individual involved in the study, and to the extent possible, the identity of those participating in the evaluation will remain confidential. Although the Associate Director and the Evaluation Specialist will know who is conducting the evaluation at the time, and office records will make it possible to link project monitors with evaluations retrospectively, the name of the project monitor will not appear on the report nor will any comments within the report be attributed to the project monitor.

Data Processing

All evaluation results will be entered into the FRONTIERS database in the Washington, D.C. office. When a sufficient number of evaluations have been completed, it will be possible to analyze the portfolio of FRONTIERS subprojects, either as a whole or by category such as country, region, topic or design type. In the long term, analyses can explore such factors as determinants of impact and relationships between utilization, process indicators and other contextual data such as family planning effort scores, contraceptive prevalence or socio-economic indicators.

In the shorter term, an evaluation update, including summary tables of indicators for all completed evaluations, will be prepared in Washington, D.C. and sent to FRONTIERS offices every six months, allowing feedback to all FRONTIERS staff and partners.

A number of evaluations, both process and impact, will be selected for external verification by the Tulane/FRONTIERS staff. The purpose of this verification is to ensure the validity and objectivity of the self-assessments. The evaluator(s) will review project documents and conduct key informant interviews, in order to assess the subproject on some or all of the indicators in the assessment form.

This type of systematic approach to the evaluation of FRONTIERS subprojects will allow Population Council staff to assess their own performance in the implementation of an OR project against a set of established standards for good OR practice. The exercise communicates to project monitors the expectations of the Washington, D.C. and regional offices, and it is expected to create greater awareness of these issues during the conduct of the studies. The impact portion of the assessment will serve to underscore the importance of getting the results utilized for the purposes of improved service delivery and in exceptional cases for changes in national policy.

Limitations of Methodology

There are a number of limitations to this approach to assessing impact of OR projects. First is the issue of attribution. It is rarely the case that an OR study alone results in a major change in service delivery or policy, and it is virtually impossible to demonstrate cause and effect in this type of impact assessment. Instead, the evaluation will attempt to demonstrate “plausible attribution,” which requires that 1) the change in service delivery take place after the intervention, and 2) the change that occurs is consistent with the results and recommendations of the OR studies.

Second, due to the three-year interval between a study’s end and the impact assessment, it may be difficult to locate some individuals who participated in the study and/or were potential utilizers of the findings. Staff turnover in local agencies, governments or Population Council offices can not be avoided, but using present project monitors to conduct evaluation will help minimize the negative effects, as they will concurrently be monitoring FP/RH projects and will thus be able to identify other suitable informants.

Third, the assessment of process and impact is qualitative in nature and thus has the limitations of any evaluation requiring qualitative judgments. Although the evaluators are expected to be objective, they must gather information from various sources and make subjective judgments in rating a specific project on a specific indicator. Verification visits confirming the findings, however, will increase the reliability of these subjective judgments.

Despite these limitations, several aspects of the process contribute to the credibility of the findings. There are no problems with sampling or lack of representativeness; the assessment

generally includes the universe of available studies. Further, the approach is systematic: a set of indicators is applied to each project using a predetermined discussion guide. In sum, even given its limitations, this evaluation methodology makes the best use of the skills of Population Council staff to provide thorough, accurate, reliable results on the impact of FRONTIERS studies.

PART B

IMPLEMENTATION

THE EVALUATION PROCEDURE

Process Assessment

The first evaluation of each subproject will deal primarily with process and contextual factors, but to the extent that any impact is visible, it should be recorded. Process information is collected primarily to better understand factors that affect utilization of results, in order to increase the impact of future studies.

The prime monitor will be responsible for completing the assessment form and preparing the evaluation report (see Attachment 1 for the Process Assessment Form). Most project monitors will know the project thoroughly and will be concurrently preparing the project final report; therefore they might complete the assessment form without the aid of outside sources of information. It may be of assistance on some indicators to review project documents to find specific examples to support a score given, but in many cases the project monitor will be sufficiently familiar with the facts of the subproject to be able to provide the necessary concrete examples without a document review.

If the prime monitor cannot complete the evaluation (for example, if he or she has left the Population Council), the Regional Associate Director should designate someone else to do so. In situations where the person completing the evaluation was not the project monitor throughout the life of the subproject, he or she may be unable to independently assess the subproject on some or all of the indicators. In such cases, the evaluator will need to rely on key informant interviews and relevant project documents. This is most likely to occur during the impact assessment and so is discussed below.

Impact Assessment

The Impact Assessment Form, Attachment 2 of this document, is used two years after the subproject has finished. FRONTIERS/Tulane staff in the Washington, D.C. office will remind project monitors and their supervisors of approaching impact evaluations. The assessment should begin 24 months after the end date, with a report submitted two months later. As a first step of the impact evaluation, to refresh his or her memory and to prepare for any key informant interviews, the project monitor should review all relevant project documents, including the final report and the process evaluation report, and document any subsequent activities to promote utilization and their effect.

The purpose of the impact evaluation is to determine what changes have occurred within the implementing or collaborating organizations as well as in other institutions or countries, as a result of the OR study. The project monitor may be aware of some but not all of these changes. Therefore, key informant interviews may help identify impact that would otherwise have been missed.

Given turnover in FRONTIERS staff, it is expected that a sizable number of the impact evaluations, as well as some of the end of project evaluations, will be conducted by a person who was not the project monitor for the study. The alternate evaluators assigned to this role by the Regional Associate Director will probably need to rely on key informants and written reports to some degree to adequately address all of the indicators.

Interviews with Key Informants

While project monitors may be able to adequately assess process independently, they will likely find a variety of viewpoints helpful in responding to some of the impact questions two years later. Ideally, the project monitor should interview key informants from each of the following categories:

- program managers and (if appropriate) providers in the service delivery organizations that stand to benefit from the OR;
- policymakers and key decisionmakers;
- donor agency staff (USAID or others such as WHO, GTZ, DFID); and
- researchers (especially Principal Investigators).

The number of key informants necessary will depend to some extent on their knowledge and memory, but in general project monitors should interview a minimum of one individual per category, while two per category will ensure consistency of information across different sources.

Experience suggests that the project monitor should begin the series of key informant interviews with the principal investigator or other researchers involved in the subproject, then continue interviewing others involved in the subproject, including staff from implementing or collaborating organizations (including high level administrators/policy makers, middle management, and where appropriate, service providers).

These interviews may be conducted in person or by phone and should take no longer than one to two weeks to complete. As with all Population Council research, informed consent should be obtained prior to conducting the interviews. Consent may be verbal or in writing.

Contact information for all key informants should be included as an appendix to the evaluation report. Direct quotes or paraphrasing of statements that support the project monitor's assessment of the subproject should be included in the evaluation report but under most circumstances they should not be attributed to a specific individual. Informants are likely to be more frank when they are assured a certain level of confidentiality.

Key informants, having different roles and perspectives in relation to the subproject, will sometimes contradict either one another or the opinions of the project monitor. In these cases, the project monitor will need to probe further in an attempt to understand the differences of opinion, and possibly contact informants from previous interviews to clarify their views. The

qualitative nature of the assessment makes explicit weighting rules impractical; the project monitor should consider the opinions of the key informants, the context and his or her own knowledge of the subproject and make the most objective assessment possible for each indicator.

THE ASSESSMENT FORM

Types of Indicators

The evaluation is based on three types of indicators:

- **Process:** the approach used in carrying out the OR projects and the quality of the research itself;
- **Impact:** evidence of change in service delivery practices or national policy related to reproductive health (RH) and family planning (FP) following the OR project; and
- **Contextual factors:** events that facilitated or hindered the conduct of the OR project and/or the utilization of results.

Fourteen process indicators, 11 impact indicators and six contextual factors have been selected for this system. They are labeled “P” for process, “I” for impact, and “C” for contextual. Thus, “P-3” refers to the third process indicator. A list of the indicators, along with a brief explanation of each, is presented below.

SCORING

The project monitor should score the study for each indicator using the following scale:

<p style="text-align: center;"><u>1</u></p> <p style="text-align: center;">Slightly or not at all</p>	<p>A score of “1” should be given when a subproject achieved up to one-third of its potential on that indicator.</p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> • A study that met only one of several objectives would receive a score of “1” on indicator P-3, “Did the study accomplish its research objectives?” • If an intervention was initially scaled up, but by the time of the impact assessment all activities tested in the intervention had been dropped, it should score a “1” on I-3, “Were the activities tested under the intervention still observable 24 months post-implementation?”
<p style="text-align: center;"><u>2</u></p> <p style="text-align: center;">Somewhat</p>	<p>“Somewhat” can be interpreted as modest achievement, or approximately one-third to two-thirds true. If a study is judged to have performed <i>adequately</i> on a given indicator but not particularly <i>well</i> (i.e. performance was less than desirable) it should score a “2”.</p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> • A study that held dissemination meetings attended by program managers and service providers but not policy makers should score a “2” on P-14, “Were the results disseminated to key audiences including policy makers, program managers, and service providers?” • If some changes made according to the recommendations of the study findings are still in place but others have been abandoned in the intervening years, the study should receive a “2” on I-3, “Were the activities tested under the intervention still observable 24 months post-implementation?”
<p style="text-align: center;"><u>3</u></p> <p style="text-align: center;">A great deal</p>	<p>A score of “3” indicates that the study achieved more than two-thirds of an indicator’s potential. Full potential will vary by indicator and subproject, but in general can be interpreted as the maximum a subproject can realistically hope to achieve on that indicator.</p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> • A study that led to a new contraceptive method (e.g. female condom) being offered in MOH clinics nationally would score a “3” on I-7, “Was there a change in policy that can be linked to the OR project?” • A score of “3” should be given for I-8, “Did the implementing organization conduct subsequent OR studies” if, following their participation in the study, gets involved additional studies and begins to use OR as a problem-solving and decision-making tool.

Descriptive Narrative in Support of the Scores

In addition to the numerical rating, the project monitor should provide a narrative explanation for the score given on each indicator. *All* scores should be supported by concrete examples, and on indicators where a subproject performed exceptionally well or poorly, project monitors should attempt to identify the reasons. These explanations give meaning to the scores, and as such are the real substance of the evaluation. Explanations should be clear and succinct and may be in the form of bulleted lists, where appropriate. If key informants were consulted, direct quotes can be used to enrich the evaluation findings, although the person interviewed should remain anonymous.

Key informants will sometimes differ in their assessment of a subproject's performance. Here the project monitor should use his or her judgment in discerning "the truth" and assigning a score, addressing the differences in the text below. Such differences in perspective can offer great insight into apparent successes or shortcomings of a study. Similarly, any exceptions to a given score should be noted in the text. For example, in one case study, evaluators discovered that service delivery improvements were still in place (Indicator I-3) at all sites except for one, where the providers had been called to military service following the outbreak of war.

Sample responses are included along with definitions for each indicator in the next section.

INDICATORS

Process Indicators

P-1: Did the implementing/collaborating organization(s) actively participate in the design of the OR project?

The design of the OR project is the formulation of the study, which includes identifying the problem, establishing the objectives, designing the intervention, and selecting a research methodology. “Active” participation involves contributing original ideas to the work, not simply attending meetings.

Example

The study was requested by ASBEF to answer questions of client flow, waiting time and client reception. ASBEF participated fully in the design of the OR project in collaboration with Population Council Staff.

Score: 3

A key informant from APROFAM said, “... In short, APROFAM staff identified the lack of CBD referrals to clinics as a problem. It was the RIGHT moment to work on this together. The heads of the two sections were at each other’s throats. Lots of tension— would there be an office of CBD within the new clinics? [The Population Council representative] served as a facilitator in this process. This was the study with the MOST participation in the design and implementation.”

Score: 3

P-2: Did the implementing/collaborating organization(s) actively participate in the implementation of the OR study?

“Active participation” indicates that the organization was involved in decision-making and played a technical role in the implementation of the study, for example hiring new staff, conducting training or analyzing results. In the case of in-house subagreements in which the implementing organization is the Population Council particular attention should be given to participation of the collaborating organization(s).

Example

With the dissolution of the Reproductive Health Unit, its training staff involved in the first study left the MOH and worked on this study as consultants for the Population Council. The MOH had given the Council approval to work in a limited area, where they conducted a training of trainers for nurses. These nurses then took over the training, resulting in substantial participation by the local MOH.

Score: 3

It was a mistake to contract people from outside [the NGO] rather than using their own personnel to do the project. It delayed the process (though it may have been more “objective”). [The NGO] had less ownership of the project.

Score: 2

P-3: Did the implementing/collaborating organization(s) participate in developing programmatic recommendations?

This indicator asks whether these organizations participated, as well as how. What evidence is there that this occurred? What form did it take: collaboration in report preparation, through formal meetings, working groups at dissemination conferences, others?

Example

The approach used to ensure participation was a series of dissemination seminars, first at the national level, then in the 10 regions, to present the results to program managers and service providers. Participants in these seminars “interpreted” the findings and developed recommendations for programmatic improvement.

Personnel at the district level responded very positively to being given the opportunity to participate directly in interpreting the results and providing input into the national and regional plan of action. They felt “valorisés.”

Score: 3

Yes, IGSS held a meeting to develop recommendations, including deciding whether it was safe to provide IUDs to women immediately following treatment for incomplete abortion.

IGSS decided it was safe and continued to offer the service, although the Population Council was not certain that this was the best decision.

Score: 3

P-4: Did the study accomplish its research objectives?

The project monitor should list the objectives from the report and indicate whether the study produced data on each objective.

Example

The purpose of this study is (1) to document the feasibility of undertaking interventions to improve the quality of care, and (2) to measure the impact of improved quality on clients' behavior.

There were two problems in accomplishing the research objectives. First, the intervention of improving quality did not take place uniformly over the five centers in the experimental group, nor was there clear documentation on what the interventions entailed. Second, although the results showed that the 5 reference centers did offer a higher quality of care than did the five control centers, the limitations in the study design do not render convincing results. (This problem is explained in detail in the text of the report.)

Score: 1

The project's objectives were to:

- *Develop a computerized MIS*
- *Implement a contraceptive distribution system*
- *Train health promoters and supervisors*
- *Develop supervision guides*
- *Identify the FP/RH needs of the migrant population*
 - *Test the use of a necklace to teach fertility awareness and natural family planning*

The study accomplished all of these objectives, although it did not lead to an increase in CPR as hoped.

Score: 3

P-5: Was the intervention implemented as planned (or with some modifications)?

Changes between the proposal and implementation of the intervention frequently occur and often are for the better. This indicator seeks to determine whether all of the activities specified in the intervention were carried out, allowing for some change in response to local realities. If not, the reviewer should identify any changes between the design and actual realization of these activities.

This indicator is not intended to penalize an organization for making modifications. Rather, it ascertains that some meaningful change was made in service delivery (that there was “something to evaluate”). When an intervention study fails to show any change in the desired outcome, there are two plausible reasons: (1) the intervention was never implemented or it was implemented so weakly that the study hardly constituted a fair test of its potential effectiveness, or (2) the intervention was fully implemented but failed to show the expected results. This indicator attempts to eliminate the first possibility by determining that the intervention was in fact implemented.

Example

The study experienced problems of contamination with the original control group. Thus they added a second control group (C-2) at the end of the study. The distributors in C-2 had similar characteristics to the original control group and had also received the routine supervision and training.

Score: 2

The primary component of the intervention was training of thana-level TFPO, ATFPOs, senior Family Welfare Visitors (FWVs), and storekeepers to determine the probable levels of contraceptive need, based on the different models being tested (volume distributed in the last month, volume distributed in the last calendar year, etc.). Conducted in November and December 1994. PIACT did a process evaluation including a mystery client to ensure that changes took place.

Score: 3

P-6: Was the study completed without delays (or other adjustments to the timeline) that would compromise the validity of the research design?

Study activities are often delayed. This indicator seeks to identify delays that affected the timing of the intervention or that could have reduced the effectiveness of certain activities (e.g., a delay in training resulted in diluting the effects of the activity; the time between intervention and final data collection had to be cut short, allowing little time for the desired change to take place). “Other adjustments to the timeline” could be an abbreviation of study activities to accommodate an approaching project end date or unfavorable season.

Example

It took almost 2 years to get the agreement of the Ministry of Health for this sensitive project. Thus, the research was only implemented in the last part of the funding year. As a result, the time between the end of the intervention and the post-test was shortened (less than 1 month).

Score: 1

The study design called for interviewing the panel of new users at 1, 6, 12, and 27 months. However, there were delays in the implementation of the intervention (related to signature of the contract between PC and MOH) and consequently in conduct of the study. Because the OR/TA II Project was coming to completion, the report on this study had to be published based on the 6-month follow-up. (The 12-month follow-up was carried out around month 17, and the 27-month follow-up is currently being done.) The length of time with which to observe contraceptive continuation was thus shortened considerably. The data collection has continued for the subsequent rounds.

Score: 2

P-7: Was continuity in key personnel maintained over the life of the OR project?

“Key personnel” are any personnel who have a decision-making role in the design or implementation of the subproject. This includes the Principal Investigator, the study coordinator, counterparts in the collaborating agencies, including key service personnel, or government officials who are active participants in implementation, as well as FRONTIERS staff.

Example

Key individuals were present in the implementing organizations. At the end of the project both CARE and MOH had personnel changes that affected continuation of the project.

Score: 1

Key individuals are still present in Le Dantec clinic ensuring continuity. Although one of the investigators left during the study, there were no adverse effects because the rest of the project personnel remained constant and were able to continue the study activities.

Score: 3

P-8: Was the study design methodologically sound (free of flaws that could have affected the final results)?

This item will be particularly pertinent during the verification visits, and should be assessed based on the methodology section of the report and (if appropriate) discussions with the researchers. Generally, the external evaluator (not a staff member of any of the participating organizations) makes an “informed decision” on this point; key informants may have less knowledge or experience to make this judgment. Most project monitors will have sufficient knowledge of the study design to be able to respond to this indicator independently, but those reviewing their assessment should pay particular attention to this indicator to minimize the possibility of bias.

Example

The study design was quasi-experimental consisting in comparison of two groups formed by the random assignment of women to experimental groups by month of childbirth. The experimental group of mothers was exposed to the intervention. The control group of mothers was exposed to the existing system of postpartum care, with some promotion of breastfeeding for reasons of child health, but virtually no information on the FP services available in the hospital.

Quantitative and qualitative data was collected from 720 women registered in the Maternity Ward. They were interviewed in the hospital during the first day after giving birth and then during a follow-up at six months postpartum. Cost data was gathered on estimates of the direct marginal costs of conducting the intervention.

Score: 3

This study was essentially a descriptive study rather than a hypothesis-driven one. Therefore, the use of qualitative methods to address questions of attitudes, norms and decision-making power was very appropriate. Quantitative methods were also appropriately used for closed-ended questions. However, there is some concern that the individuals sampled were selected only from among women who had undergone a tubal ligation or were currently using Norplant. This design excludes an important segment of the population – those women who choose not to use long-term methods. If the objective is to provide information to develop IEC materials, it is important to have information, not only from women who make the unusual decision to select these long-term methods (given the cultural context), but also to gather data from women who elect NOT to use long-term methods

Score: 2

P-9: Was the research design feasible in the local context?

“Feasible” should be interpreted as “reasonable” or “manageable”: a design that could be repeated without unduly draining financial or human resources. “Local context” includes not only program-related factors but also sociocultural or political factors, among others.

Example

The research design was feasible because the criteria for mothers participating in the study were tied to practical issues. Only women who gave birth between Monday and Friday were included; women who lived outside San Pedro Sula were excluded to reduce the costs of follow-up; women whose infants died during the hospital stay or whose infant was identified at high risk were excluded.

Score: 3

The design was logistically feasible, but given the sensitive nature of the subject (post-abortion care), it took a lot of advocacy time and effort before the project was seen as acceptable for the local context.

Possibly, with other organizations or individuals, and with more time dedicated to developing the relationship between the two NGOs and conducting formative research, but not as it was carried out.

Score: 1

P-10: Did the implementing/collaborating organizations judge the OR technical assistance to be useful and provided in a collegial manner?

To qualify for a score of 3, both elements must be positive. If, for example, the advice was technically sound, but counterparts reacted negatively to the manner in which it was provided (e.g., in an offensive or condescending way, “imposed upon them”), then the study should receive a score of 2 on this indicator.

Example

Although APROFAM felt that the TA was methodologically sound, they felt that it was not always given in a “collegial” manner (e.g., that the PC advisor tried to impose her ideas on the others; that discussions became tense when problems arose.)

Score: 2

P-11: Were results of the OR study judged to be credible/valid in the local context?

This indicator refers to the judgment of stakeholders (policy makers, researchers, donors, program managers). It is assumed that utilization of results would be limited if stakeholders seriously questioned the validity of the results.

Example

Results had to be interpreted cautiously. Contamination forced a change in the original design and it was difficult to determine substantial differences between the pre- and post-tests in certain areas.

Score: 1

Although the results of the study did not show the intervention to be particularly successful, the new MOH authorities knew results and were very interested in disseminating them, as well as revising the algorithm and continuing to test it through OR.

Score: 3

P-12: Was the research relevant for the national program?

Relevance is defined based on the perceptions of the same stakeholders listed above. Relevant research addresses a priority problem of the program, whether a national program of the MOH or a more local program of an NGO.

Example

The topic of improving the quality of services is a major part of the PFPN program. Thus, a study to measure whether the model clinics do indeed provide higher quality service and whether that improved service results in higher continuation rates is highly relevant to local concerns.

Score: 3

P-13: Were the results disseminated to key audiences, including policy makers, program managers, service providers, and donors?

All studies involve dissemination of results. This indicator seeks to determine whether the dissemination strategies used were effective in reaching the target audience. “Key audiences” are those in a position to act on the results (e.g., policy makers, key decision-makers or service providers in implementing/collaborating agencies, donor agency staff). In addition, dissemination efforts may reach other interested parties (e.g., students at the local university, members of the international FP/RH community), but the indicator refers only to those in a position to act upon the results.

Example

The results were disseminated to senior MOH officials through a series of small meetings with handouts consisting of a brief summary and visuals used in the presentation. Other MOH staff at the central and regional levels heard about the study from a paragraph included in the MOH newsletter and through a special notice to supervisors of district facilities. Senior program managers and representatives of five family planning facilities attended the one-day seminar held in the capital to discuss the study findings and compile policy recommendations. Research staff have made presentations before meetings of health providers and health advocacy groups. The study was mentioned in a radio news report and two newspaper articles. More than 300 copies of a 10-page summary of the Final Report were distributed to key policymakers, program managers, service providers at public and private facilities, researchers, libraries, and journalists.

Score: 3

P-14: Are the results readily available in written form?

This refers primarily to the existence of a document on the key findings of the study that is well presented (of professional quality) and is locally available in sufficient quantity. This document may be presented in a variety of media (e.g. website, CD-ROM) in addition to print. Ideally, results should be available in various formats appropriate to the intended audience: final reports and journal articles for donors and the academic RH community, summaries or research briefs for decision makers and program managers. The project monitor should provide the full citation of all products describing the research, as well as the language of each.

Example

The final report is available in English only, making it of limited use locally. (However, as a result of the evaluation, it will be summarized and translated into Spanish.) In addition, although the organization has the report, the key informant said, they would have difficulty locating it if anyone wanted it.

Score: 2

- *A summary report was prepared in French: Mané B, Dieng T, Faye O. Tapsoba P, Diadhiou F. Introduction des soins obstétricaux d'urgence et de la planification familiale pour les patientes présentant des complications liées à un avortement incomplet. JHPIEGO, Population Council, CEFOREP, CGO, décembre, 1998.*
- *A paper has been prepared for submission to a peer-reviewed journal (not yet submitted at the time of the evaluation): Dieng T, Mané B, Faye O, Tapsoba P, Mbengue CA, Diadhiou F. La planification familiale (PF) pour les patientes présentant des complications liées à un avortement. CEFOREP.*
- *A summary of the project was included in the Africa OR/TAII Project Final report. Postabortion Care. Africa OR/TA II Project Final Report. 1998, p. 61-67.*
- *A report of the literature review of abortion in Senegal was prepared: Camara CM, Cissé L. Revue de la littérature sur les avortements à risque au Sénégal. JHPIEGO, Population Council, CEFOREP, CGO, avril 1998.*
- *A "better practices summary" was prepared (English): "Better Practices in Reproductive and Child Health: Introducing improved Postabortion Care in Maternity Clinics of University Teaching Hospitals in Burkina Faso and Senegal, p. 1-3."*

Score: 3

Impact Indicators

I-1: Did the results indicate that the intervention was effective (i.e. that it improved service delivery in the areas identified by the study)?

OR studies generally either test one or more interventions, or they evaluate changes resulting from interventions already implemented. If all studies found the intervention under study to be effective, there would be no need to conduct the research. This indicator asks whether the intervention tested was evaluated to be successful in improving front-line service delivery (e.g., increase in utilization of services, improved quality of services). Negative results can also be instructive, but they would not influence service delivery except to discontinue an ineffective strategy (see I-2).

Example

The intervention proved to be only somewhat effective. With the algorithm, MOH administrators and providers could see RH as more than simply family planning, and more easily integrate RH services. The training increased knowledge and abilities necessary for providing systematic integrated services. However, as few providers use the algorithm as a job aid, lost opportunities were not reduced to the extent hoped.

Score: 2

The average length of stay was reduced from approximately 3 days to 1 day (subsequently reducing costs), counseling for family planning is done in a more systematic manner (while the woman is still in the hospital), and then reinforced in a follow-up visit one week later. Sterile procedures that were part of the PAC regimen have been adopted in other areas of the hospital and contributed to a reduction in infection. In pilot sites, a greater variety of family planning services are now available (none existed in the military hospital prior to the study).

Score: 3

I-2: Did the implementing/collaborating organization(s) “act on” the results (i.e., continue to implement the activities tested in the OR study after its completion if effective or not implement/discontinue this activity if ineffective)?

“Acting on the results” consists of implementing the actual services of the intervention or the activities to support those services (e.g., training courses, development of service delivery guidelines, changes in allocation of personnel, production and testing of IEC materials, supervision, monitoring) if the intervention was effective, or not implementing or discontinuing these services and activities if the intervention proved not to be effective.

Example

The PNPf of the MOH, with support from USAID (in 4 regions) and UNFPA (in others):

- *Installed 14 Centre de Reference (one per region, plus four additional ones) during the period from 1995-97;*
- *Purchased equipment needed for service delivery;*
- *elaborated a protocol of norms and procedures jointly financed with UNFPA;*
- *developed and produced IEC materials;*
- *provided training in FP service delivery and counseling at the national and regional level ;*
- *developed a guide on contraceptive logistics management (with assistance from FPLM)*

At the district level:

- *managers empowered by availability of data to make decisions at local level, not wait for direction from central level;*
- *managers used the OR results to develop their work plans for 1996;*
- *district level authorities agreed to standardize the price of FP services;*
- *many districts adopted the COPE methodology;*
- *new supervisory mechanisms were introduced, with support from USAID in 4 regions*

Score: 3

ASBEF gave more specific instructions to nurses' aides to avoid improper screening of clients. They reorganized the layout and redefined tasks, improving patient flow. Interventions to cut down waiting time were implemented and midwives took a more active role in supervising the nurses' aides. This study was one of the motivating factors leading ASBEF to construct a new facility. COPE was also conducted as a result of this study.

Score: 3

**I-3: (If the intervention was effective and continued after the study)
Were the activities tested under the intervention still observable 24
months post-implementation?**

“Activities tested under the intervention” are those specific items mentioned in connection with the previous indicator. Where only some of the original activities are observable, the study should receive only a partial score on this indicator. In the case of an improvement that has been in effect less than 24 months, this indicator is not applicable.

Example

Activities were in effect for at least five years following the study. Data is not available for the subsequent years, due to difficulty contacting AGROSALUD staff. However, health posts continue to offer some family planning services and referrals.

Score: 3

I-4: If the intervention was effective and continued after the study, was the intervention scaled up by the original implementing/collaborating organization in the same country?

Most OR studies are conducted in a specific geographical area. “Scaling up” can be defined as implementing the intervention activities in additional geographical areas. It can but does not necessarily have to refer to expansion to the national level. Some OR projects (such as Situation Analysis) do not test interventions, and thus these indicators would be “not applicable.”

Example

Le Dantec/CEFOREP is part of the University and its role is to pilot strategies. The Ministry of Health is expanding PAC into 5 sites outside Dakar and there is demand to expand it even further. Health Posts have requested that PAC be integrated into their structure although there is much debate as to the appropriateness of that setting for PAC. Further scaling up is projected after assessing feasibility outside Dakar.

Score: 3

I-5: If the intervention was effective and continued after the study, was the intervention adopted by another organization within the same country?

An organization that did not participate in the OR study “adopts” the intervention by implementing its primary components (see I-1).

Example A (Score: 3)

Norplant has become one of the methods of choice offered by the Ministry of Health and ASBEF is now offering Norplant as well. Tubal ligation is now offered in several other regions and is coordinated by AVSC.

Example B (Score: 2)

Some NGOs such as Project Hope used the results and built on the intervention, involving health promoters. MotherCare adopted certain parts of the intervention, in particular the training methodology and began to include self-esteem as part of their trainings.

I-6: Was the intervention replicated in another country?

There must be some evidence that the original intervention contributed to the activities being carried out in the other country (e.g., program managers from other countries visited the project site and subsequently adopted similar strategies).

Example

Senegal replicated the study that was first conducted in Burkina Faso. However, they are now assisting Burkina Faso to provide technical assistance for conducting similar OR studies in Guinea, Madagascar, Côte d'Ivoire and Haiti. Mali and Benin await funding to replicate the intervention.

Score: 3

Bolivia and Dominican Republic asked for technical support in project implementation in those countries.

Score: 2

I-7: Was there a change in policy that can be linked to the OR project?

Policy change refers legislation or other official changes that potentially affect service delivery; for example, authorization for the sale of the pill by non-medical personnel. Changes at the local or program level should be noted, as well as changes in Ministry of Health or other national policy.

Example

The study did not attempt to change national policy and there was no dissemination beyond the NGO, but there was a change in AGROSALUD policies, to continue to offer the services tested in the study within AGROSALUD's constellation of services.

Score: 3

I-8: Did the implementing/collaborating organization conduct subsequent OR studies?

“Subsequent OR studies” refers specifically to research activities that test interventions. OR studies do NOT include actions or program activities, such as training and materials production (described in I-2), without a research component, or research for other purposes (e.g., the DHS, epidemiological research).

Example

This study provided the opportunity for CEFORP to establish a research division, thus adding research capacity to its training activities. They are currently conducting other research studies on adolescent and men’s health.

Score: 3

I-9: Did the implementing/collaborating organization conduct subsequent OR studies without the Population Council?

This indicator is included to reflect whether the organization has sufficient capacity to conduct these types of activities as a result of the previous OR experience (although in some cases they may have had this capacity already). Similarly, the organization may have the capacity but no opportunity to conduct these types of studies. If any other organization provided technical assistance, the project monitor should note the name of the organization and type of technical assistance.

Example

This initial OR study introduced the idea of operations research to ASBEF. They have gone on to conduct some small internal OR projects, particularly to assess waiting time. Additionally, they have conducted a small OR project on CBD for family planning.

Score: 2

I-10: Did the original donor fund new program activities based on the results of the OR study?

New program activities are those activities tested in the intervention that were not already funded by the donor.

Example

USAID funded the purchase of equipment for facilities in four regions of Senegal, in addition to training of personnel in counseling and FP service delivery, and IEC materials.

Score: 3

I-11: Did other donors provide new or expanded funding based on results of the OR studies?

“Other donors” are those donor agencies that did not contribute financial support to the original OR project but subsequently funded the initiation or expansion of program activities—specifically, service delivery or support activities, including training, production of IEC materials, construction or renovations of facilities and purchase of supplies and equipment, among others. It does not include funding of additional research only.

Example

UNFPA provided funding for training and the purchase of equipment for family planning in the remaining six regions of Senegal.

Score: 3

Contextual Factors

C-1: Were there other factors (not mentioned above) that facilitated the conduct of the research project?

Situations that helped in the completion of the activity might be: strong research capacity in the counterpart organization, powerful local person or donor intent on getting answers, good relationship between researchers and program staff, or others.

Example A

One of the counselors to the Minister of Health sat on the technical advisory committee, assuring that information flowed to the Minister and helping to gain his agreement to conduct the study.

Example B

The widespread interest in the topic of quality of care (among local program managers in Senegal) and the intense interest among the international FP/RH community to have empirical evidence on the link between improved services and client behavior.

C-2: Were there other factors (not mentioned above) that facilitated the utilization of results?

This refers to situations that encourage the translation of the results into programmatic actions at the field level. Examples: dissemination of results coincides with the planning cycle for a new program initiative or strategy, the intervention is a good match with the organization, a committed individual continues to provide TA or promote the intervention beyond the project end.

Example

UNFPA was involved in interim results dissemination and exposed to the intervention early. Thus, by the time the project was over, they had already incorporated it into their new 5-year plan.

C-3: Were there other factors (not mentioned above) that hindered the conduct of the research project?

Events beyond the control of the researchers and their collaborators in local organizations may impede a study's implementation. Such factors range from contraceptive stock-outs, inter-organizational or interpersonal conflicts, or financial difficulties, to political changes, civil unrest or natural disasters.

Example A

Physicians in the regions were displeased that the head of the PNPf was a sociologist, not a person with medical training. This created some resistance at first toward participating in the study, but eventually they willingly participated.

Example B

Physicians from the military hospital trained in the pilot were called to duty when the war in neighboring Guinea Bissau broke out, leaving the site without trained providers. Nurse midwives were subsequently trained to perform MVA, reinstating and expanding the capacity to provide the intervention.

Example C

The data collection was contingent on the completion of the model clinics and related improvements (e.g., training of personnel). Delays in this area appear to have hindered progress on data collection. Moreover, the uneven timing in completing these improvements meant that the job was not really done to the same degree in all five model clinics by the time the data collection for the panel of new clients began.

C-4: Were there other factors (not mentioned above) that hindered the utilization of results?

Situations that limit the incorporation of research results into policy or programs may also be beyond the control of researchers and their collaborators. Some examples are: the intended population opposes the intervention (e.g., believes that contraceptives cause sterility), local authorities for political reasons veto a proposed initiative or change in service delivery, or the health system is restructured.

Example

APROFAM couldn't scale up fast enough to meet full demand, because not all promoters met the eligibility criteria. The medical community is always a bit negative toward having a product like this in the hands of promoters.

C-5: Did USAID use the data from the OR study for a specific purpose? (Explain)

Is there any evidence that USAID used the data for its own purposes? For example, results might be incorporated into the MAQ initiative or used as part of the R4 process, for funding decisions, RFA design or for miscellaneous strategic planning.

Example A

CERPOD is the logical organization to continue this activity, but they suffer from an overburdened staff with too many activities to respond to.

Example B

Because USAID/Dakar needed reliable data for its R4 reporting to Washington, the Mission strongly supported this round of SA data collection.

C-6: Did the study include an assessment of the costs of the intervention?

Any data collected on the cost of the intervention, primarily for the purpose of cost-effectiveness analysis, should be mentioned. This indicator is used for informational purposes only, since all OR studies do not necessarily need an assessment of cost.

Example

The study included a cost analysis of the ongoing needs of the reproductive health unit.

THE EVALUATION REPORT

For each assessment conducted, either process or impact, the project monitor will produce an evaluation report. The report will consist of a cover sheet briefly summarizing the methodology of the study and the evaluation (described in detail below), and the completed assessment form. The background data of each study is contained in the ACCESS database in the FRONTIERS/DC office and will be used to generate a cover sheet for each subproject, which will be sent to the project monitor two months before the report is due. In addition to completing the assessment form, the project monitor should verify that the information on the cover sheet is accurate, and fill out the methodology, limitations and comments sections.

Each report will first be reviewed by the regional Associate Director, and then submitted to the Evaluation Specialist in Washington, D.C. A list of key informants contacted should be included as an appendix.

Project title:

If the title of the subproject does not match the title on the final report, the evaluation report should use the subproject title.

Agreement Number:

This is the agreement number between the Population Council and the implementing organization.

Dates of project:

Official start and end dates, as noted in documentation to USAID.

Implementing organization:

The agency that signs a contractual agreement to carry out the subproject.

Collaborating organization(s):

Other agency(ies) that play a technical role in the design or implementation of the subproject. (Note: attending a dissemination conference only does not qualify as playing a technical role.)

Project summary:

Brief description of the primary objectives, research design, and results of the study or activity (maximum: 250 words).

Methodology:

Describe the type of evaluation (i.e. end of project, impact), the role of the evaluator in the implementation of the subproject and the methods of data collection (such as self-report, key informant interviews, written documents).

Limitations:

There are some limitations inherent to the methodology, presented in Section II, above. The Limitations section of the report refers to OTHER limitations specific to the evaluation. These limitations may vary somewhat with the role of the evaluator and methods of data collection, but some examples are: the project monitor was assigned to evaluate a study he or she did not previously work on, a key informant was judged not to be credible, or insufficient data was available to assess the project on a certain indicator

Comments:

This section should be used to communicate anything of importance not addressed elsewhere. The project monitor may wish add relevant facts related to process or impact that do not fit into the assessment form, or discuss any of the information presented. He or she may also use this space to comment on the evaluation process itself.